



TITLE:

輸出入統計データベースシステムの設計

AUTHOR(S):

宮崎, 耕

CITATION:

宮崎, 耕. 輸出入統計データベースシステムの設計. 経済論叢 1999, 163(1): 64-74

ISSUE DATE:

1999-01

URL:

<https://doi.org/10.14989/45258>

RIGHT:

經濟論叢

第163巻 第1号
定道 宏教授記念號

献 辞	渡 邊 尚	
インターネットを利用した 遠隔合同ゼミナールの実現	布 上 康 夫	1
パーソナルウェアの概念と機能	松 本 良 治	37
デジタルビデオ編集システム	広 田 雅 彦	51
輸出入統計データベースシステムの設計	宮 崎 耕	64
インターネット時代における グループウェア・メール	高 井 才 明	75
デジタルユニバーシティへの第一歩	細 井 真 人	99
国際会議とインターネット	中 村 素 典	112

定道 宏 教授 略歴・著作目録

平成11年1月

京 都 大 学 經 済 學 會

輸出入統計データベースシステムの設計¹⁾

宮 崎 耕

はじめに

「輸出入統計データベースシステム」は，京都大学経済学部の定道 宏教授を中心とするグループによって整備が進められている「世界経済総合データベースシステム」を構成するサブデータベースシステムの1つであり，わが国の貿易統計を，相手国別，品目別の月次データ（数量および価額）で格納する経済統計データベースシステムである。

本論文は，「輸出入統計データベースシステム」をモデルケースとして，経済統計データベースシステムの設計について論じるものである。

I 世界経済総合データベースシステム

「世界経済総合データベースシステム」²⁾（WEDS：World Economic Database System）は，わが国と世界主要諸国の国民経済計算統計，外国貿易統計などの主要経済指標を格納し，国内外の研究者に学術情報サービスとして公開することを目指して，京都大学経済学部内に本拠地をおく「世界経済総合データベース構築グループ」³⁾によって1988年度から構築が進められているユニークな経済統計データベースシステムである。

1) 本研究は，1996年度同志社大学学術奨励研究——研究課題「WWWとDBMSの連携によるマクロ経済統計データベースサービスシステムの開発」——及び，文部省科学研究費補助金——研究成果公開促進費（データベース）[申請番号67（平成7年度），申請番号86（平成8年度），申請番号94（平成9年度）]——の交付を受けて実施した研究の成果の一部である。

2) 定道 [1990] [1993 a] [1993 b] [1993 c] [1994]，宮崎 [1995 a] [1995 b] 参照。

3) 代表者 定道 宏 京都大学大学院経済学研究科教授。

WEDS は、単体のデータベースシステムではなく、複数のサブシステムによって構成される。現在までに、SNA⁴⁾ 統計データを収録する「世界マクロ経済統計データベースシステム」⁵⁾、計量経済モデルを収録する「計量経済モデルデータベースシステム」⁶⁾ および、わが国の貿易統計を収録する「日本貿易統計データベースシステム」⁷⁾ が構築された。

本論文でモデルケースとする「輸出入統計データベースシステム」は、これらのサブシステムのうち、「日本貿易統計データベースシステム」を代替するものとして現在開発中のものである。

II 大蔵省貿易統計

「輸出入統計データベースシステム」には、大蔵省関税局が発行する「大蔵省貿易統計」の統計データが収録されている。

この統計は、227の国と地域に対する、輸出7846品目、輸入9920品目についての数量および価額の月次データであり、その件数は例えば1996年12月の場合1ヶ月で輸出7万3684件、輸入4万293件の計11万3977件にのぼる。「輸出入統計データベースシステム」に当初収録を予定している1978年1月～1996年12月の輸出入データ全体では、約2000万レコード、データ本体のみの容量で約13億バイトに及ぶ膨大なものとなる⁸⁾。(表1)

III 世界経済総合データベースシステムのデータ構造

WEDS のデータ構造については、時系列データベースのデータ構造について論じた定道 [1991] で提示され、現在までに構築されたサブシステムはいず

4) System of National Accounts.

5) 宮崎 [1990] [1995 b] 参照。

6) 定道 [1993 c]; Sugiura [1988] 参照。

7) 中原 [1992] [1993] 参照。

8) 輸出1441万1186レコード、輸入578万7975レコードの計2019万9161レコード、1レコードあたりのデータ量が65バイトであるため、計13億1294万5465バイト＝約1.2ギガバイトとなる。

表1 「大蔵省貿易統計」のデータ件数 (単位: 件)

年	輸 出	輸 入	計
1978年	654,823	195,508	850,331
1979年	664,838	214,301	879,139
1980年	728,602	210,429	939,031
1981年	744,538	211,102	955,640
1982年	705,271	206,568	911,839
1983年	672,734	200,122	872,856
1984年	680,770	209,774	890,544
1985年	684,482	216,925	901,407
1986年	651,355	228,270	879,625
1987年	617,614	252,180	869,794
1988年	827,637	342,863	1,170,500
1989年	840,224	369,327	1,209,551
1990年	874,203	387,838	1,262,041
1991年	874,950	391,880	1,266,830
1992年	882,501	392,540	1,275,041
1993年	843,809	394,971	1,238,780
1994年	819,583	421,644	1,241,227
1995年	807,064	454,170	1,261,234
1996年	836,188	487,563	1,323,751
計	14,411,186	5,787,975	20,199,161

れもここに示めされたデータ構造に準拠して設計されている⁹⁾。その要点はつぎのとおりである。

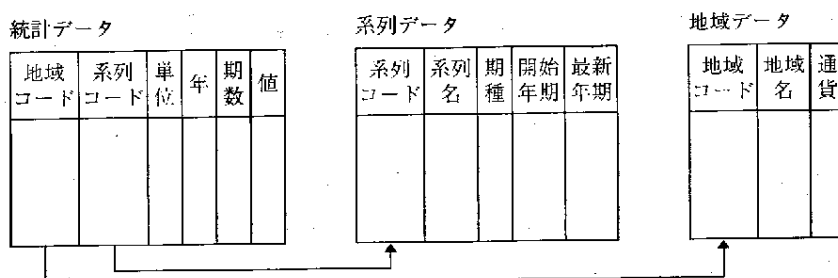
リレーショナルデータベースでは、正規化理論をよりどころとして、リレーションの構成などの基本的なデータ構造が決定される。正規化理論にしたがえば、時系列データベースは、基本的につぎのような3つのリレーションで構成される。(図1)

また、SQL¹⁰⁾ をもちいてこのデータベースから任意の時系列データを検索・抽出することが可能であり、このデータ構造で理論的には問題なく機能す

9) WEDS のサブシステムは、いずれもリレーショナルデータベース管理システム (RDBMS: Relational DataBase Management System) のもとに構築されている。

10) Structured Query Language: リレーショナルデータベースに対するデータ操作の標準言語。

図1 リレーショナルデータベース理論に基づく時系列データベースの構造



ることが確認できる。

しかしながら、時系列の経済統計データは、次数（項目数）は少ないものの、濃度（レコード件数）が極端に大きいという特徴をもつ。例えば、貿易統計の場合、ソースデータの項目数はわずか11であるのに対して、レコード件数は、すでにしめしたように18年間で約2000万レコードにもなる。（表2）

表2 「大蔵省貿易統計」のデータ項目

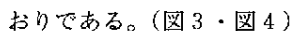
輸出入 区分	年	初 月	当 月	品目 コード	国 コード	第一 数量	第一 単位	第二 数量	第二 単位	価額 千円
-----------	---	--------	--------	-----------	----------	----------	----------	----------	----------	----------

このように膨大な件数のレコードを格納するリレーションに対してデータ操作を行うことは極めて不効率であり、実用的なデータベースサービスを提供することは困難である。

そこで、WEDS では、理論データモデルのリレーションをマスターとして、月次・四半期・年次などの期種別のデータ操作用リレーションをそれぞれ作成し、ビューとしてではなく実表としてデータベースに格納する。（図2）このデータ構造を採用することで、データ操作時における実質的な操作対象レコード数を圧縮し、効率的なデータ検索・抽出を可能としているのである。

このデータ構造を用いて設計された「世界マクロ経済統計データベースシステム」および「日本貿易統計データベースシステム」のデータ構造はつぎのと

年次リレーション



「日本貿易統計データベースシステム」は1992年にパイロットシステムが稼働して以来、試験的なサービスを提供してきた。

この間、利用者から出された改善要求や利用実態を総合すると——①年次データを扱う場合でも、なお操作対象レコードの数が多すぎる。②利用者の大半は品目別ではなく、より粗い種類別のデータを所望している。③主要国を除

図3 世界マクロ経済統計データベースシステムのデータ構造

国 コード	系列 コード	単位	初 年	期種 コード	1 期 値	2 期 値	3 期 値	4 期 値	5 期 値	...	56 期 値	57 期 値	58 期 値	59 期 値	60 期 値

図4 日本貿易統計データベースシステムのデータ構造

月次データ

国 コード	品目 コード	第一 単位	第二 単位	1月 第一 数量	1月 第二 数量	1月 価額	2月 第一 数量	2月 第二 数量	2月 価額	...	11月 第一 数量	11月 第二 数量	11月 価額	12月 第一 数量	12月 第二 数量	12月 価額

年次データ

国 コード	品目 コード	年	第一 単位	第二 単位	第一 数量	第二 数量	価額

けば、利用者の多くは、国別よりも粗いブロック別のデータを所望している。

——という3つの課題が明らかとなった。

①については、現状のデータ構造では年次データとして取り出してもレコード数は12分の1にしかならず、18年間では単純計算で約170万件のレコードが操作対象となっていることに最大の問題点があると考えられる。しかしながら、仮に10年分を1レコードとするデータ構造に変更しても、レコード数は約17万件であり不十分である¹⁰⁾。

表3 大蔵省が発表する
「地域別貿易統計データ」の地域区分

ア ジ ア	西 欧
中華人民共和国	ド イ ツ
大 韓 民 国	イ ギ リ ス
台 湾	フ ラ ン ス
香 港	オ ラ ン ダ
シンガポール	イ タ リ ア
タ イ	ベ ル ギ ー
マレーシア	ス イ ス
インドネシア	ス エ ー デ ン
フィリピン	アイルランド
ヴィエトナム	ス ペ イ ン
インド	(EU)
(アジア NIEs)	中東欧・ロシア等
(ASEAN)	ロ シ ア
大 洋 州	中 東
オーストラリア	アラブ首長国連邦
ニュー・ジーランド	サウディアラビア
北 米	ク ウ ェ イ ト
ア メ リ カ	イ ラ ン
カ ナ ダ	ア フ リ カ
中 南 米	南アフリカ共和国
ブ ラ ジ ル	
メ キ シ コ	
チ リ	総 額

表4 大蔵省が発表する
「主要商品貿易統計データ」の商品区分

食 料 品	電 気 機 器
繊維及び同製品	重 電 機 器
合成繊維織物	映 像 機 器
化 学 製 品	(テレビ受像機)
有機化合物	(VTR)
プラスチック	(テレビカメラ)
非金属鉱物製品	音 響 機 器
陶 磁 器	(ラジオ受信機)
金属及び同製品	音響・映像機器の部分品
鉄	通 信 機
非 鉄 金 属	半導体等電子部品
金 属 製 品	(IC)
一 般 機 械	電 気 計 測 機 器
原 動 機	輸 送 用 機 器
事務用機器	白 動 車
(自動データ処理機械)	(乗用車)
金属加工機械	(バス・トラック)
建設用・鉱山用機械	自動車の部分品
加熱用・冷却用機器	二 輪 自 動 車
ポンプ・遠心分離機	船 舶
荷 役 機 械	そ の 他
精 密 機 器	紙 類・紙 製 品
科学光学機器	ゴムタイヤ・チューブ
(複写機)	写真用・映画用材料
時 計	レコード・テープ類
計 総 額	

②, ③については, 例えばつぎのような集計データが所望されている。(表3・表4)

- 11) ユーザがリレーション内の全レコードダウンロードして分析を行うことを想定するならば, 例えば分析ツールとして広く利用されている「Microsoft Excel」の1シートで扱えるデータの最大行数が65536行であることなどにも配慮する必要がある。

このような集計値は、SQL での問い合わせを工夫することによって抽出可能なものではあるが、実用的なサービスを提供するためには、データ構造の抜本的見直しが必要である。

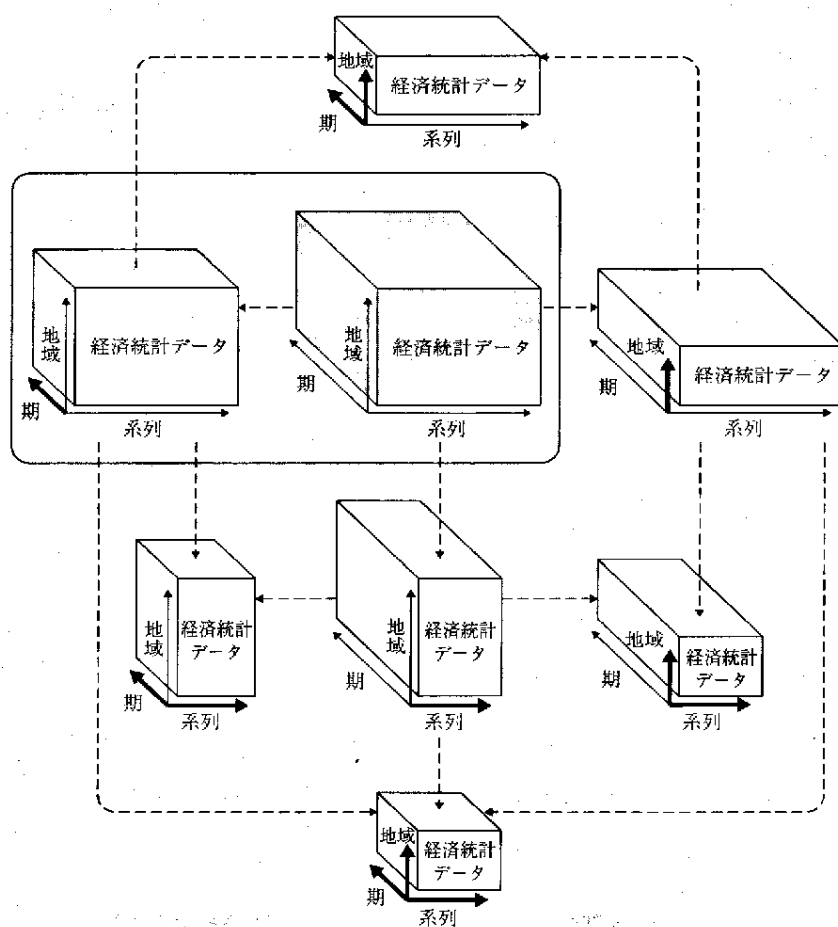
「日本貿易統計データベースシステム」に関する以上のような課題に対処するため、「輸出入統計データベースシステム」の設計においては、定道 [1991] でしめされたデータ構造の設計モデルを拡張する。すなわち、レコード数の圧縮を、時系列方向だけではなく品目や国についても行い、利用者のニーズに即したより多くの集計済みデータを収容したりレションを実表としてデータベースに格納するのである。(図5)

お わ り に

本論文では、経済統計データを期(時間)軸、系列軸、地域軸の3次元の構造をもつデータとしてとらえ、各軸について検索・抽出などのデータ操作時における操作対象レコード数を圧縮することが、実用的な経済統計データベースシステムを設計する上でのポイントとなることを指摘し、その設計モデルを提示した。(図5)

この設計モデルに基づいて、高デマンドの集計処理を事前に施した集計表を原始データとは別個の実表に収納してデータベースに格納すれば、データベース利用時における実質的な操作対象レコード数を削減し、検索効率を高めるとともに、オンデマンドで CPU に高い負荷がかかる集計処理が繰り返し実行されることを回避し、システム全体のパフォーマンスを向上させることが可能である。

データベースに、内部で生成可能なデータを重複して格納することは、データの完全性や記憶スペースの合理的利用の観点からは回避すべきである。しかしながら、経済統計データには、取引データを扱う業務データベースとは異なり——①データの更新周期が長い、②データの更新は基本的にレコードの追加のみであり、収録済みのデータを変更することはない、③データ更新の即時性

図5 経済統計データベースの設計モデル¹²⁾

12) □内は、定道 [1991] で提示された部分をしめす。

性は必ずしも要求されないため、大規模なデータ処理が必要とされる集計表の作成は計算機資源に余裕のあるタイミングに適宜バッチ処理で実行することができる。——といった特性があり、システム全体からみれば、集計値の格納は必ずしも冗長であるとはいえない。

なお、本論文で提示した設計に基づく「輸出入統計データベースシステム」の構築および実装については別稿にてしめすこととする。

参考文献

- 定道 宏 [1989] 「統計データベースの普及とサービス体制」『経済経営研究年報』第39号 (I・II)。
- 定道 宏 [1990] 「世界経済総合データベースの通信ネットワーク」『経済経営研究年報』第40号 (II)。
- 定道 宏 [1991] 「SQL と時系列データベースについて」『国民経済雑誌』第164巻第5号。
- 定道 宏, 宮崎 耕, 中原昭宏 [1993 a] 「世界経済総合データベースシステムの開発」『オフィス・オートメーション』Vol. 14, No. 2。
- 定道 宏 [1993 b] 「BBS による統計 DB の国際利用」『学術月報』日本学術振興会, 1993年9月。
- 定道 宏, 布上康夫 [1993 c] 「BBS による計量モデルベースの国際的利用」『オフィス・オートメーション』Vol. 14, No. 3-4。
- 定道 宏 [1994] 「世界経済総合データベースの国際的利用」『第2回関西情報関連学会連合大会講演論文集』。
- Sugiura, I. [1988] "Distributed Control of Economic Data Bases and Models" in *National Income and Economic Progress*, ed. by D. Ironmonger et al., GBR, Macmillan Press.
- 中原昭宏 [1990] 「BBS ソフトと DBMS ソフトを結合したデータベース利用通信システムの開発」『経済経営研究年報』第40号 (II)。
- 中原昭宏 [1992] 「日本貿易統計データベースシステム」『経済経営研究年報』第42号。
- 中原昭宏 [1993] 「BBS による日本貿易統計データベースシステムの利用」『経済経営研究年報』第43号。
- 中原昭宏 [1995] 「本邦国民経済計算統計データベースのユーザインターフェース」

『オフィス・オートメーション』Vol. 16, No. 2。

宮崎 耕 [1990] 「世界マクロ経済統計データベースシステム」『経済経営研究年報』第40号(Ⅱ)。

宮崎 耕 [1995 a] 「世界経済総合データベースシステムの構築」『国民経済雑誌』第171巻第3号。

宮崎 耕 [1995 b] 『ネットワーク環境における情報システムの研究』神戸大学経済経営研究所 研究叢書45。